DOCUMENT RESUME

ED 349 125 PS 020 827

TITLE Alternatives for Measuring Performance. Guidebook 4. Schools That Work: The Research Advantage.

INSTITUTION National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.; North Central Regional Educational Lab., Oak Brook, IL.; Public Broadcasting Service, Washington, DC. PBS Elementary/Secondary Service.

SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.

PUB DATE 91

CONTRACT RP91002007

NOTE 43p.; Part of an eight-program videoconference. For other guidebooks in the series, see PS 020 824-831.

AVAILABLE FROM Publications Department, North Central Regional Educational Laboratory, 1900 Spring Road, Suite 300, Oak Brook, IL 60521-1480 (Videotape and guidebook, $25.00).

PUB TYPE Guides - Non-Classroom Use (055) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Check Lists; Competency Based Education; *Educational Practices; Educational Research; Educational Testing; Educational Trends; Elementary Secondary Education; *Evaluation Research; Research and Development; School Community Relationship; *Student Evaluation; *Theory Practice Relationship

IDENTIFIERS *Performance Based Evaluation; *Video Teleconferencing

ABSTRACT

A series of eight videoconferences was designed to help educational professionals improve schools by applying knowledge gained from research. During the videoconference programs, which are transmitted by satellite to facilities with receiving capacity, viewers can interact by telephone with the program presenters. Each program covers a specific topic and each has a "content partner" (a learning center, institute, teachers' association or other group) that supplies expertise in the program area. The format of the program involves the presentation of two stories depicting classrooms or communities that have successfully integrated research with practice, and a question-and-answer session with the studio audience and remote viewers. In this document, a review of research on the nature of student assessment and the effects of assessment on educational practice is followed by an overview of the fourth program in the series, which covers the topic of alternative methods of student assessment. In this program, the first story describes the use of a wide range of assessment practices in third- and fourth-grade classes in an elementary school in Los Angeles. The second story illustrates the ways in which a Wisconsin school district is changing assessment practices through the active involvement of the school and community. A series of follow-up activities relevant to the topic of student assessment is described. Additional materials include biographical information on videoconference participants, a list of resources on student assessment, a glossary of terms, and a reference list of 23 items. (BC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED349125

SCHOOLS THAT WORK
The Research Advantage

GUIDEBOOK

4

Alternatives for Measuring Performance

PS 020827

An Eight-Program Videoconference

presented by the
North Central Regional Educational Laboratory
and the PBS Elementary/Secondary Service
in partnership with the
National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

BEST COPY AVAILABLE

NCREL

The North Central Regional Educational Laboratory (NCREL) is one of ten federally supported regional educational laboratories in the country and serves the states of Illinois, Indiana, Iowa, Michigan, Minnesota, Ohio, and Wisconsin. NCREL works with education professionals in the region to support school restructuring to promote learning for all students, especially those most at risk of academic failure in the most rural and urban settings.

NCREL was created in 1984 by seven chief state school officers and 19 deans of education from colleges and universities. It is governed by a 29-member board of directors composed of the seven chief state school officers and representatives from higher education, business, labor, agriculture, local and intermediate educational agencies, and communities.

The Lab's efforts to respond to regional needs are focused around five major program areas which direct our efforts to those in greatest need and allow us to concentrate our resources in areas likely to result in the greatest impact: curriculum, instruction, and assessment; early childhood and family education; professional development; rural education; and urban education.

As an educational laboratory, NCREL assists schools, families, and communities in identifying and addressing the needs of our children and students and in redesigning schools so that they have a greater capacity to create learning environments appropriate to, and supportive of, the full diversity of student needs and circumstances. We believe a growing body of expertise embodied in theory, research, and practice is pointing toward ways in which the nation can fundamentally rethink, or restructure, its schools.

PBS

The PBS Elementary/Secondary Service (ESS), through and on behalf of PBS member stations, serves the instructional needs of students and teachers nationwide. ESS distributes educational programming and related materials for K-12 classroom use; promotes the use of PBS's primetime programming as a curricular resource; provides professional development for educators; and serves as a national advocate for the use of video and related technologies for learning.

The PBS Adult Learning Service (ALS) offers college-credit television courses through local partnerships of public television stations and colleges. Since 1981, more than 1,500 colleges, in cooperation with 300 public television stations, have enrolled more than one million students in ALS-distributed courses. In August 1988, ALS launched the PBS Adult Learning Satellite Service (ALSS) as a direct satellite service for higher education, offering a wide variety of programming.

National Center for Research on Evaluation, Standards, and Student Testing (CRESST)

The National Center for Research on Evaluation, Standards, and Student Testing (CRESST) has committed itself to the research and development of improved approaches to assessment at all levels of education. Funded by the Office of Educational Research and Improvement, and administratively housed at the UCLA Center for the Study of Evaluation, CRESST is a unique partnership of leading researchers and practitioners from across the country.

Five major goals constitute the current CRESST mission:

1. *Provide* leadership to improve assessment policy and practice at the national, state, and local levels
2. *Improve* the quality, sensitivity, and fairness of student performance assessments
3. *Improve* the validity of models and indicators for judging the quality of school
4. *Improve* understanding of assessment development, implementation, and effects as they occur in school practice
5. *Improve* understanding of assessment policy and its contribution to educational improvement

Economically disadvantaged and minority students are a continuing focus of CRESST research and development efforts.

The CRESST team is composed of researchers from the following institutions: UCLA Center for the Study of Evaluation; University of Colorado; University of Chicago; University of Pittsburgh; the RAND Corporation; and University of California, Santa Barbara. CRESST also benefits from cooperative relationships with practitioner groups, professional organizations, business and industry, the military, and individual consultants.

CRESST maintains and distributes over 330 technical reports to customers across the United States and abroad. Additionally, CRESST's quarterly newsletter, *CRESST Line*, reaches a base of over 5,000 educational professionals and organizations throughout the United States. Distributed free, *CRESST Line* contains current information on CRESST research projects, activities, and news.

For more information we invite you to contact us at (310) 206-1532. Or write to:

CRESST/UCLA Graduate School of Education
145 Moore Hall
405 Hilgard Avenue
Los Angeles, California 90024-1522

What is a Videoconference?

The North Central Regional Educational Laboratory (NCREL) and the PBS Elementary/Secondary Service have been producing videoconferences together since 1987. Via satellite, videoconferencing allows the exchange of information across town or across the country. A growing number of organizations use this delivery system for staff development, corporate meetings, training sessions, and entertainment events.

Videoconferencing works like this: An "uplink" station transmits the live program to a communications satellite which then beams it to "downlink" sites. Any facility with a satellite dish and the proper receiving capacity can be a downlink site. When combined with a telephone hookup, videoconferencing becomes "interactive" with one-way video (viewers see the presenters on monitors although presenters cannot see the viewers) and two-way audio (viewers can call in and speak to the presenters). Another way to include an interactive component is to have a viewer call-in period following the live satellite transmission.

In 1990, NCREL and PBS produced a nine-part staff development videoconference series, *Restructuring to Promote Learning in America's Schools*. Although not geared specifically toward staff development, *Schools That Work: The Research Advantage* has been designed to help viewers develop the kinds of successful schools that ensure a successful future for all of us.

©1991 North Central Regional Educational Laboratory
1900 Spring Road, Suite 300, Oak Brook, IL 60521
Jeri Nowakowski, Executive Director

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means without permission.

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement (OERI), Department of Education, under Contract Number RP91002007. The content of this publication does not necessarily reflect the views of OERI, the Department of Education, or any other agency of the U.S. Government.

Acknowledgments

NCREL Video Series

Content Partner

National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
University of California, Los Angeles

Project Team

Beau Fly Jones, Project Director
Randy A. Knuth, Program Associate
Steve Baxendale, Program Assistant
Judson Hixson, Program Director
Ernest Rose, Evaluator
Mary Fran Tilton, Secretary
June Yang, Video Production
Annette Mallory, Video Production
Louis Ciancio, Video Production

Video Development

General Learning Video

Larry Walcoff, Senior Producer
Nick Kolias, Senior Camera Crew
Jack O'Neal, Senior Camera Crew

Uplink Station

WTTW Chicago

Jack Ginay, Director
Steve Norman, Production Manager
Linda Weseman, Production Coordinator

Guidebook

Written by

Ronald J. Dietel
Director of Communications, CRESST
University of California
Los Angeles, California

Joan L. Herman
Associate Director, CRESST
University of California
Los Angeles, California

Randy A. Knuth
North Central Regional Educational Laboratory

With Contributions from

Tammy Chaney, Consultant
Indiana University
Bloomington, Indiana

Linda Christensen
Director of Instruction
Verona Area High School
Verona, Wisconsin

Guidebook continued

Charlotte Higuchi, Teacher
Farmdale Elementary School
Los Angeles, California

Diana Varner
Learning Resource Coordinator
Verona Area High School
Verona, Wisconsin

Formative Field Review Team

Mary Goosby, Principal
Louis Armstrong School
Chicago, Illinois

Tom Stefonek, Director, Bureau for Educational Statistics and Testing
Wisconsin Department of Public Instruction
Madison, Wisconsin

Elaine McEwan, Asst. Superintendent for Instruction
West Chicago Elementary Dist. 33
West Chicago, Illinois

Marlene Szymanski, Principal
Manirre School
Chicago, Illinois

F. Mayo Sanders, Principal
Vistula Park School
Fremont, Indiana

Judy Zito, Principal
Tioga Elementary School
Bensenville, Illinois

Printing and Dissemination

Editing and Print Production

Marianne Kroeger, Editor
Stephanie Merrick, Editor
Lenaya Raack, Editor

Sharon Sherman, Desktop Publisher
Cheryl May, Production
Geri Goelz, Production

Dissemination Team

Beverly Walker, Urban
Lawrence Friedman, Rural
Linda Kunesh, Early Childhood

Gordon Hoke, Policy
Margaret Gould, Dissemination

Table of Contents

SECTION 1: Overview of the Videoconference Series 1
- Topics and Schedule ... 1
- Audience ... 1
- Videoconference Programs ... 2
- Videoconference Objectives ... 3
- The National Education Goals ... 3
- Rationale for the Videoconference Program Topics ... 4
- Format of Each Videoconference Program ... 5

SECTION 2: What Does Research Say About Assessment? 6
- Purposes of Assessment ... 6
- Effects of Traditional Tests ... 7
- Characteristics of Good Assessment ... 8
- Criteria for Valid Performance-Based Assessments ... 9
- Findings from Cognitive Psychology ... 10
- Important Trends in Assessment ... 11
- Trends Stemming from the Behavioral to Cognitive Shift ... 12
- Issues of Equity and Assessment ... 13
- Social Organization and Assessment ... 14

SECTION 3: Overview of Program 4, "Alternatives for Measuring Performance" 15
- Program Objectives ... 15
- Program Outline ... 15
- Overarching Message ... 16
- About Story 1: Assessment Strategies in the Classroom ... 17
- About Story 2: The Community and Verona Area High School ... 18
- Viewers' Guide for Program 4, "Alternatives for Measuring Performance" ... 19

SECTION 4: Follow-up Activities for Participants 20
- Activities for Teachers ... 20
- Activities for Schools, Parents, and Community Members ... 21
- Checklist for Excellence in Assessment ... 23

SECTION 5: Biographical Information 25
- Videoconference Participants ... 25

SECTION 6: Important Assessment Resources 28

SECTION 7: Glossary 30

SECTION 8: References 33

Product Order Form 35

SECTION 1

Overview of the Videoconference Series

The President of the United States and the National Governors Association recently set forth six goals to be achieved by the year 2000 (see page 3). *Schools That Work: The Research Advantage* is designed to provide educational research for schools and communities as they reflect upon and seek to implement these national education goals. A recent Gallup poll shows that most Americans strongly support the national goals. Each goal was rated as a high priority by at least 84% of those polled. However, at least 47% felt that reaching these goals by the year 2000 is unlikely.

The North Central Regional Educational Laboratory and the PBS Elementary/Secondary Service are committed to restructuring schools to promote learning. We believe that every child can learn. We also believe that it is in our nation's utmost interest to develop world-class schools that achieve these goals, for the sake of our children.

Scattered throughout America are schools that work. These schools involve teachers, students, administrators, parents, community members, and businesses. They illustrate how the application of learning research can lead to improved education and to the achievement of the national goals. These examples give us hope that we all can make the difference. This videoconference is designed to help viewers develop schools that work, schools that aim to meet the national goals using the research advantage.

Topics and Schedule

1991 (4:00-5:00 pm ET)	**1992** (4:00-5:00 pm ET)
Reading—October 10	Early Childhood Education—February 13
Mathematics—October 24	Drug-Free Schools—March 12
Science—November 14	School-to-Work Transition—April 2
Assessment—November 21	Interagency Collaboration—May 7

Audience

Schools That Work: The Research Advantage is designed primarily for teachers, administrators, and curriculum specialists, as well as staff of state education and intermediate service agencies. However, its fast-paced, informative style will appeal to school board members, parents, policymakers, Chapter I staff, teacher educators and their students, test makers, staff developers, and researchers interested in school improvement.

Videoconference Programs

Videoconference 1

Reading ▪ October 10, 1991 ▪ Children as Strategic Readers

Content Partner: Center for the Study of Reading (CSR), University of Illinois, Urbana-Champaign

Videoconference 2

Mathematics ▪ October 24, 1991 ▪ Children as Problem Solvers

Content Partners: National Center for Research in Mathematical Sciences Education, University of Wisconsin, Madison, and the National Council of Teachers of Mathematics

Videoconference 3

Science ▪ November 14, 1991 ▪ Children as Explorers

Content Partners: Institute for Research on Teaching, Michigan State University, and Project 2061, American Association for the Advancement of Science (AAAS)

Videoconference 4

Assessment ▪ November 21, 1991 ▪ Alternatives for Measuring Performance

Content Partner: National Center for Research on Evaluation, Standards, and Student Testing (CRESST), UCLA

Videoconference 5

Early Childhood Education ▪ February 13, 1992 ▪ Meeting Children's Needs

Content Partners: National Association for the Education of Young Children (NAEYC) and the National Association of Early Childhood Specialists in State Departments of Education

Videoconference 6

Drug-Free Schools ▪ March 12, 1992 ▪ Preparing Students for Drug-Free Lifestyles

Content Partner: Midwest Regional Center for Drug-Free Schools and Communities

Videoconference 7

School-to-Work Transition ▪ April 2, 1992 ▪ Preparing Students for Work in the 21st Century

Content Partners: Institute on Education and the Economy, Teachers College, Columbia University, and the U.S. Chamber of Commerce

Videoconference 8

Interagency Collaboration ▪ May 7, 1992 ▪ Integrating Community Services

Content Partners: Study Group on School-Linked Integrated Services sponsored by the U.S. Department of Education and the Department of Health and Human Services

Videoconference Objectives

- To provide examples of how research empowers teachers, students, and whole communities
- To promote meaningful and self-directed learning for all students
- To focus national attention on issues of equity and excellence as schools and communities implement the national goals
- To demonstrate alternative ways that educators, parents, policymakers, and community members can work together to redesign schools for America's children
- To provide high-quality programming with up-to-date research for professional development libraries for educators
- To encourage schools and communities to reflect on the national education goals, well informed by the best of research and practice

The National Education Goals

Goal 1: By the year 2000, all children in America will start school ready to learn.

Goal 2: By the year 2000, the high school graduation rate will increase to at least 90 percent.

Goal 3: By the year 2000, American students will leave grades four, eight, and twelve having demonstrated competency in challenging subject matter including English, mathematics, science, history, and geography; and every school in America will ensure that all students learn to use their minds well, so they may be prepared for responsible citizenship, further learning, and productive employment in our modern economy.

Goal 4: By the year 2000, U.S. students will be first in the world in science and mathematics achievement.

Goal 5: By the year 2000, every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship.

Goal 6: By the year 2000, every school in America will be free of drugs and violence and will offer a disciplined environment conducive to learning.

Rationale for the Videoconference Program Topics

The topics for this series were selected in part because they represent major areas of educational research and in part because they are all relevant to the national education goals. Although each topic area in the series involves different departments and school staff, it is important to understand that schools ultimately are part of one system. What happens in one area of schooling influences what happens in another area. This videoconference series presents two sets of thematically related programs.

- The theme connecting the first four programs (Reading, Mathematics, Science, Assessment) is that research in these areas forms the vision of learning and teaching that drives curriculum, instruction, and assessment. Further, many common threads exist in each research base such as teaching for understanding and collaborative learning.
- The theme connecting the last four programs (Early Childhood Education, Drug Education, School-to-Work Transition, Interagency Collaboration) is that traditionally these topics have been considered somehow outside the core of schooling—both conceptually and politically. Our intent is to demonstrate that they are vital for the achievement of all children. Further, the collaborative learning and higher order thinking, together with their assessment, needed in the subject disciplines are also vital to early childhood education, preparing students for drug-free schools, and the work environment.

Accordingly, NCREL and PBS strongly recommend that schools and community members watch all eight programs to develop a common language and common objectives. Indeed, we hope that schools and community members will use this series as a vehicle for debate and reflection about school reform and implementing the national goals.

Format of Each Videoconference Program

Live Satellite Transmission (1 hour)

For educational research to have a positive impact on schooling, it must be supported by effective classroom practices. With this in mind, the satellite transmission portion of each videoconference will focus on two "stories" depicting actual classrooms or communities that have successfully integrated research into practice. Each story will be followed by a question-and-answer session with the live studio audience. Remote viewers will be able to call in at any time during the program and give their questions to the telephone operator. The moderator will ask selected questions to the presenters throughout the program. Questions not answered on the air may be answered by phone or in print.

The first story will demonstrate effective classroom practices. The second will portray the rich teaching and learning environment created through community involvement in schooling. Each program in the videoconference series draws upon the best of research and practice to address critical questions that schools and communities need to resolve if they are to implement the national goals. Specifically, each program will illustrate the following:

1. How research has changed an individual teacher and the lives of his/her students
2. How schools have used research in the community by involving parents or community members in planning, instruction, or assessment or have enabled teachers and students to engage in learning outside the traditional boundaries of the school

Participant Call-In (30 minutes)

Immediately following the satellite transmission of each videoconference, participants will be given the opportunity to "call in" with comments and questions for videoconference presenters. This segment does not include live video transmission. However, callers will speak directly to the presenters. Because this segment provides one-to-one contact between presenters and participants, it is the key to the interactive nature of the conferences.

SECTION 2

What Does Research Say About Assessment?

Assessment may be defined as "any method used to better understand the current knowledge that a student possesses." This implies that assessment can be as simple as a teacher's subjective judgment based on a single observation of student performance, or as complex as a five-hour standardized test. The idea of current knowledge implies that what a student knows is always changing and that we can make judgments about student achievement through comparisons over a period of time. Assessment may affect decisions about grades, advancement, placement, instructional needs, and curriculum.

Purposes of Assessment

The reasons why we assess vary considerably across many groups of people within the educational community.

Who Needs To Assess?	Purposes of Assessment
Policymakers use assessment to:	• Set standards • Focus on goals • Monitor the quality of education • Reward/sanction various practices • Formulate policies • Direct resources including personnel and money • Determine effects of tests
Administrators and school planners use assessment to:	• Monitor program effectiveness • Identify program strengths and weaknesses • Designate program priorities • Assess alternatives • Plan and improve programs
Teachers and administrators use assessment to:	• Make grouping decisions • Perform individual diagnosis and prescription • Monitor student progress • Carry out curriculum evaluation and refinement • Provide mastery/promotion/grading and other feedback • Motivate students • Determine grades
Parents and students use assessment to:	• Gauge student progress • Assess student strengths and weaknesses • Determine school accountability • Make informed educational and career decisions

Effects of Traditional Tests

Billions of dollars are spent each year on education, yet there is widespread dissatisfaction with our educational system among educators, parents, policymakers, and the business community. Efforts to reform and restructure schools have focused attention on the role of assessment in school improvement. After years of increases in the quantity of formalized testing and the consequences of poor test scores, many educators have begun to strongly criticize the measures used to monitor student performance and evaluate programs. They claim that traditional measures fail to assess significant learning outcomes and thereby undermine curriculum, instruction, and policy decisions.

The higher the stakes, the greater the pressure that is placed on teachers and administrators to devote more and more time to prepare students to do well on the tests. As a consequence, narrowly focused tests that emphasize recall have led to a similar narrowing of the curriculum and emphasis on rote memorization of facts with little opportunity to practice higher-order thinking skills. The timed nature of the tests and their format of one right answer has led teachers to give students practice in responding to artificially short texts and selecting the best answer rather than inventing their own questions or answers. When teachers teach to traditional tests by providing daily skill instruction in formats that closely resemble tests, their instructional practices are both ineffective and potentially detrimental due to their reliance on outmoded theories of learning and instruction.

Characteristics of Good Assessment

Good assessment information provides accurate estimates of student performance and enables teachers or other decisionmakers to make appropriate decisions. The **concept of test validity** captures these essential characteristics and the extent that an assessment actually measures what it is intended to measure, and permits appropriate generalizations about students' skills and abilities. For example, a ten-item addition/subtraction test might be administered to a student who answers nine items correctly. If the test is valid we can safely generalize that the student will likely do as well on similar items not included on the test. The results of a good test or assessment, in short, represent something beyond how students perform on a certain task or a particular set of items; they represent how a student performs on the objective which those items were intended to assess.

Measurement experts agree that test validity is tied to the purposes for which an assessment is used. Thus, a test might be valid for one purpose but inappropriate for other purposes. For example, our mathematics test might be appropriate for assessing students' mastery of addition and subtraction facts but inappropriate for identifying students who are gifted in mathematics. Evidence of validity needs to be gathered for each purpose for which an assessment is used.

A second important characteristic of good assessment information is its **consistency,** or reliability. Will the assessment results for this person or class be similar if they are gathered at some other time or under different circumstances or if they are scored by different raters? For example, if you ask someone what his/her age is on three separate occasions and in three different locations and the answer is the same each time, then that information is considered reliable. In the context of performance-based and open-ended assessment, inter-rater reliability also is essential; it requires that independent raters give the same scores to a given student response.

Other characteristics of good assessment for classroom purposes:

- The content of the tests (the knowledge and skills assessed) should match the teacher's educational objectives and instructional emphases.
- The test items should represent the full range of knowledge and skills that are the primary targets of instruction.
- Expectations for student performance should be clear.
- The assessment should be free of extraneous factors which unnecessarily confuse or inadvertently cue student responses. (For example, unclear directions and contorted questions may confuse a student and confound his/her ability to demonstrate the skills which are intended for assessment. A math item that requires reading skill will inhibit the performance of students who lack adequate skills for comprehension.)

Researchers at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) are developing an expanded set of validity criteria for performance-based, large-scale assessments. Assessment researchers Bob Linn, Eva Baker, and Steve Dunbar have identified eight criteria that performance-based assessments should meet in order to be considered valid.

Criteria for Valid Performance-Based Assessments

Criteria	Ask Yourself
Consequences	Does using an assessment lead to intended consequences or does it produce unintended consequences, such as teaching to the test? For example, minimum competency testing was intended to improve instruction and the quality of learning for students; however, its actual effects too often were otherwise (a shallow drill and kill curriculum for remedial students).
Fairness	Does the assessment enable students from all cultural backgrounds to demonstrate their skills, or does it unfairly disadvantage some students?
Transfer and generalizability	Do the results of the assessment generalize to other problems and other situations? Do they adequately represent students' performance in a given domain?
Cognitive complexity	Do the assessments adequately assess higher levels of understanding and complex thinking? We cannot assume that performance-based assessments will test a higher level of student understanding because they appear to do so. Such assumptions require empirical evidence.
Content quality	Are the tasks selected to measure a given content area worth the time and effort of students and raters?
Content coverage	Do the assessments enable adequate content coverage?
Meaningfulness	Are the assessment tasks meaningful to students and do they motivate them to perform their best?
Cost and efficiency	Has attention been given to the efficiency of the data collection designs and scoring procedures? (Performance-based assessments are by nature labor-intensive.)

Multiple-choice measures have provided a reliable and easy-to-score means of assessing student outcomes. In addition, considerable test theory and statistical techniques were developed to support their development and use. Although there is now great excitement about performance-based assessment, we still know relatively little about methods for designing and validating such assessments. CRESST is one of many organizations and schools researching the promises and realities of such assessments.

Findings from Cognitive Psychology

Analysis of Traditional Views

Methods of assessment are determined by our beliefs about learning. According to early theories of learning, complex higher-order skills had to be acquired bit-by-bit by breaking learning down into a series of prerequisite skill, a building-blocks-of-knowledge approach. It was assumed incorrectly that after basic skills had been learned by rote, they could be assembled into complex understandings and insight. However, evidence from contemporary cognitive psychology indicates that all learning requires that the learner think and actively construct evolving mental models.

From today's cognitive perspective, meaningful learning is reflective, constructive, and self-regulated. People are seen not as mere recorders of factual information but as creators of their own unique knowledge structures. To know something is not just to have received information but to have interpreted it and related it to other knowledge one already has. In addition, we now recognize the importance of knowing not just how to perform, but also when to perform and how to adapt that performance to new situations. Thus, the presence or absence of discrete bits of information—which is typically the focus of traditional multiple-choice tests—is not of primary importance in the assessment of meaningful learning. Rather, what is important is how and whether students organize, structure, and use that information in context to solve complex problems.

Cognitive Psychology

Contrary to past views of learning, cognitive psychology suggests that learning is not linear, but that it proceeds in many directions at once and at an uneven pace. Conceptual learning is not something to be delayed until a particular age or until all the basic facts have been mastered. People of all ages and ability levels constantly use and refine concepts. Furthermore, there is tremendous variety in the modes and speed with which people acquire knowledge, in the attention and memory capabilities they can apply to knowledge acquisition and performance, and in the ways in which they can demonstrate the personal meaning they have created.

Current evidence about the nature of learning makes it apparent that instruction which strongly emphasizes structured drill and practice on discrete, factual knowledge does students a major disservice. Learning isolated facts and skills is more difficult without meaningful ways to organize the information and make it easy to remember. Also, applying those skills later to solve real-world problems becomes a separate and more difficult task. Because some students have had such trouble mastering decontextualized "basics," they are rarely given the opportunity to use and develop higher-order thinking skills.

Recent studies of the integration of learning and motivation also have highlighted the importance of affective and metacognitive skills in learning. For example, recent research suggests that poor thinkers and problem solvers differ from good ones not so much in the particular skills they possess as in their failure to use them in certain tasks. Acquisition of knowledge skills is not sufficient to make one into a competent thinker or problem solver. People also need to acquire the disposition to use the skills and strategies, as well as the knowledge of when and how to apply them. These are appropriate targets of assessment.

The role of the social context of learning in shaping higher-order cognitive abilities and dispositions has also received attention over the past several years. It has been noted that real-life problems often require people to work together as a group in problem-solving situations, yet most traditional instruction and assessment have involved independent rather than small group work. Now, however, it is postulated that groups facilitate learning in several ways: modeling effective thinking strategies, scaffolding complicated performances, providing mutual constructive feedback, and valuing the elements of critical thought. Group assessments, thus, can be important.

Important Trends in Assessment

Since the influence of testing on curriculum and instruction is now widely acknowledged, educators, policymakers, and others are turning to alternative assessment methods as a tool for educational reform. The movement away from traditional, multiple-choice tests to alternative assessments—variously called authentic assessment or performance assessment—has included a wide variety of strategies such as open-ended questions, exhibits, demonstrations, hands-on execution of experiments, computer simulations, writing in many disciplines, and portfolios of student work over time. These terms and assessment strategies have led the quest for more meaningful assessments which better capture the significant outcomes we want students to achieve and better match the kinds of tasks which they will need to accomplish in order to assure their future success.

Trends Stemming from the Behavioral to Cognitive Shift

Emphasis of Assessment	Behavioral Views	Cognitive Views
View of learner	Passive, responding to environment	Active, constructing knowledge
Scope of assessment	Discrete, isolated skills	Integrated and cross-disciplinary
Beliefs about knowing and being skilled	Accumulation of isolated facts and skills	Application and use of knowledge
Emphasis of instruction and assessment	Delivering maximally effective materials	Attention to metacognition, motivation, self-determination
Characteristics of assessment	Paper-pencil, objective, multiple-choice, short answer	Authentic assessments on contextualized problems that are relevant and meaningful, emphasize higher-level thinking, do not have a single correct answer, have public standards known in advance, and are not speeded
Frequency of assessment	Single occasion	Samples over time (portfolios) which provide basis for assessment by teacher, students, and parents
Who is assessed	Individual assessment	Assessment of group process skills on collaborative tasks which focus on distributions over averages
Use of technology for administration and scoring	Machine-scored bubble sheets	High-tech applications such as computer-adaptive testing, expert systems, and simulated environments
What is assessed	Single attribute of learner	Multidimensional assessment that recognizes the variety of human abilities and talents, malleability of student ability, and that IQ is not fixed

There is continuing faith in the value of assessment for stimulating and supporting school improvement and instructional reform at national, state, and local levels. In summary, while the terms may be diverse, several common threads link alternative assessments:

- Students are involved in setting goals and criteria for assessment.
- Students perform, create, produce, or do something.
- Tasks require students to use higher-level thinking and/or problem solving skills.

- Tasks often provide measures of metacognitive skills and attitudes, collaborative skills and intrapersonal skills as well as the more usual intellectual products.
- Assessment tasks measure meaningful instructional activities.
- Tasks often are contextualized in real-world applications.
- Student responses are scored according to specified criteria, known in advance, which define standards for good performance.

Issues of Equity and Assessment

While assessment has the potential to improve learning for all students, historically it has acted as a barrier rather than a bridge to educational opportunity. Assessments have been used to label students and put them in dead end tracks. Traditional tests have been soundly criticized as biased and unfair to minority students. And, the assessment of language minority students has been particularly problematic.

A key point regarding equity as applied to performance-based assessment is made by Yale Professor Emeritus Edmund Gordon. "We begin with the conviction that it is desirable that attention be given to questions of equity early in the development of an assessment process rather than as an add-on near the end of such work....The task then is to find assessment probes (test items) which measure the same criterion from contexts and perspectives which reflect the life space and values of the learner."

Robert Linn says, "The criterion of equity needs to be applied to any assessment. It is a mistake to assume that shifting from standardized tests to performance-based assessments will eliminate concerns about biases against racial/ethnic minorities or that such a shift will necessarily lead to equality of performance.

"Although many at-risk students come to school deficient in prior knowledge that is important to school achievement, teachers and schools can make a substantial difference through the construction of assessments that take into account the vast diversity of today's student populations. Gaps in performance among groups exist because of differences in familiarity, exposure, and motivation of the subjects being assessed. Substantial changes in instructional strategy and resource allocation are required to give students adequate preparation for complex, time-consuming, open-ended assessments. Providing training and support for teachers to move in these directions is essential.

"Questions of fairness arise not only in the selection of performance tasks but in the scoring of responses. As Stiggins has stated, it is critical that the scoring procedures are designed to assure that 'performance ratings reflect the examinee's true capabilities and are not a function of the perceptions and biases of the persons evaluating the performance.' The same could be said regarding the perceptions and biases of the persons *creating* the test. The training and calibrating of raters is critical in this regard."

Social Organization and Assessment

What we know about performance-based assessment is limited and there are many issues yet to be resolved. We do know that approaches which encourage new assessment methods need the broad-based support of the community and school administration. Like any change in schools, changes in assessment practices will require:

- Strong leadership support
- Staff development and training
- Teacher ownership
- Continuing follow-up and support for change through coaching and mentoring
- Environments that support experimentation and risk-taking

As schools move toward more performance-based assessment, they also will need to come to some resolution on a number of issues, among them that performance-based assessments:

- Require more time to develop
- Cost more
- May limit content coverage
- Require a shift in teaching practices
- Require substantial time for administration
- Lack a network of colleagues for sharing and developing
- Require new methods of aggregating and reporting data
- Require new viewpoints about how to use for comparative purposes

SECTION 3

Overview of Program 4: Alternatives for Measuring Performance

Program Objectives

- To provide the major findings from research and practice in assessment
- To provide examples of excellence at classroom, school, and community levels
- To raise the kinds of questions that might be discussed in school and community forums

Program Outline

- The program opens with comments from committed individuals that serve as a preview of key ideas.
- Moderator Judson Hixson, and Beau Fly Jones, Project Director for the videoconference series, both from the North Central Regional Educational Laboratory, introduce the videoconference series and Program 4.
- The first story illustrates how research is enabling Charlotte Higuchi to integrate fair, performance-based assessment at every level of her instruction.
- Researchers Eva Baker from CRESST and Ed Gordon from Yale discuss the major findings from research reflected in Charlotte Higuchi's classroom approach.
- Hixson moderates questions and comments from the audience.
- The second story illustrates how the Verona, Wisconsin, school district is undergoing substantial change in assessment practices through the active involvement of the school and community.
- Diana Varner, Learning Resource Coordinator, and Patrick Sweeney, community resident, discuss how the changes in their district reflect research on learning and assessment.
- Hixson moderates questions and comments from the audience.
- Jones asks selected studio audience members to describe how the stories and discussion in this videoconference may influence their thinking.

Overarching Message

1. The new view that learning is a constructive, reflective, self-regulated process requires authentic, performance-based assessments that take into account the whole child.
2. Meaningful assessment plays a critical role at every level and aspect of instruction.
3. School and community collaborations—using principles from assessment research—represent a powerful strategy for developing the assessment approaches best suited for the community.

About Story 1: Assessment Strategies in the Classroom

To Charlotte Higuchi, a 3rd and 4th grade teacher at Farmdale Elementary School in Los Angeles, assessment and instruction are inseparable aspects of teaching. Ms. Higuchi has developed a wide array of assessment strategies to both determine how well her students are doing and to make instructional decisions. Although she uses standardized tests scores, they are but one small indicator among many that helps her develop a "snapshot" of progress.

Ms. Higuchi captures evidence of her students' performance in a variety of portfolio artifacts. Students place their best work into portfolios and keep journals about their classroom performance and assessment. Ms. Higuchi uses videotape to record students' performances on tasks as well as parent-teacher conferences. She also keeps a journal of her classroom instruction and each student's performance. She has developed systematic rating procedures to assess various outcomes. This collection of data helps her to paint a rich picture of class performance and allows her to develop a rich understanding of the relationships between her teaching and student learning.

In this story you will see the assessment strategies that Ms. Higuchi uses which include:

Individual Assessment

At frequent intervals during the school year, Ms. Higuchi visits one-on-one with each student. In these conferences the students discuss past progress and do a performance task (e.g., in reading or math problem solving) which Ms. Higuchi rates. These conferences are audiotaped or videotaped for later analysis and may end up in the portfolio.

Group Assessment

Working collaboratively is a major goal in Ms. Higuchi's class. Her students work together on many problem-solving activities. Similar to the individual assessment, groups meet with the teacher to analyze their progress and to perform a collaborative task which is recorded and rated.

Parent/Student/Teacher Conferences

Parental involvement is crucial to the sucess of students. Ms. Higuchi meets regularly with the student and his or her parents to discuss progress and to show evidence of performance. The teacher models assessment strategies which the parents are encouraged to use with their children at home. Parents develop a sense of responsibility from this process and become committed to their child's learning.

Whole Group Assessment

During class Ms. Higuchi is continually asking questions and "reading" the responses of her students to check for understanding. Based on these observations she makes instructional decisions.

About Story 2: The Community and Verona Area Schools

The Verona Area Schools in Wisconsin are in the midst of a major restructuring effort. About four years ago, concerned members of the school board realized that Verona had no plan for the future. At the same time a handful of teachers and administrators were developing an educational vision and plan for change called PACIS (Personalized, Authentic, Community-Based, Individualized, Student-Centered). Although innovative and reflecting many of the values of systemic reform, PACIS was eventually voted down by the school board because of its cost, and because of the lack of a shared vision in the district of the goals, purposes, and processes of education within which PACIS would operate. These events led to the start of a strategic planning process that involves representatives of every level of the school and community.

Within the strategic planning process, community residents work side by side with school personnel on "action teams." These teams debate and work out action plans for achieving eight goals that arose from community/school meetings and retreats. During the retreats, members of the community participated on panels that were interviewed by action teams to find out what educational outcomes were important to them. For example, members of the panel were asked, "From your perspective, what are the most important attributes and abilities required to be a functioning human being?" Answers ranged from definitions of citizenship to technical ability. Out of these panels arose what the Verona district calls learner outcomes which were organized into "significant spheres of living." These spheres are oriented toward human attributes and describe dispositions to create knowledge, collaborate, value diversity, act with consciousness, and so on.

The action teams were charged with defining what each of the eight goals meant, and what actions should be taken to achieve them. The first issue dealt with by an action team was assessment. Assessment was seen as a critical place to start because of the emerging philosophy that assessment, instruction, and curriculum cannot and should not be separated. The action plans specified how students, teachers, and administrators might be evaluated in an authentic, meaningful manner. These recommendations lead to discussions about authentic tasks and how the community might involve itself both as a resource for the tasks and as assessors of performance.

In this story, you will see the members of the Verona school district and community working through issues of authentic assessment and how this issue affects the larger picture of school reform. We will visit the panels of community members presenting their definitions of "functioning human beings." A look at the authentic assessment action team will illustrate how the community is shaping Verona's educational future. Students will be seen working on authentic tasks and receiving authentic assessment at a hospital blood analysis center. By using a systematic performance evaluation technique, other "assessors," and video technology, a high school literature teacher will show how she assesses group discussions. Processes of teacher and administrator assessments will be illustrated. Artifacts of assessment, including portfolios, will provide examples of how students can be authentically assessed. And throughout the story you will hear community members and key school personnel reflect about authentic assessment and the school change process.

Viewers' Guide for Program 4: Alternatives for Measuring Performance

Things to look for during the program...	Spaces for taking notes
How do the teachers and researchers define authentic assessment? How do their definitions compare to my school?	
List the full range of assessment strategies that Charlotte Higuchi uses.	
Who participates in assessment, as shown in these stories? What roles does each person take?	
How do the teachers and schools integrate performance-based assessment with traditional forms such as standardized tests?	
How are the community members of Verona involved in assessment redesign?	
How do the teachers in this program determine what to assess and what criteria to use?	
How do the approaches presented in this program address the national goals?	

SECTION 4

Follow-Up Activities for Participants

Activities for Teachers

The examples of excellence in this program clearly show that in successful schools, teaching is a multidimensional activity. One of the most powerful of these dimensions is that of "teacher as researcher." Not only do teachers need to use research in their practice, they need to participate in "action" research in which they are always engaging in investigation and striving for improved learning. The key to action research is to pose a question or goal, and then design actions and evaluate progress in a systematic, cyclical fashion as the means are carried out. Below are four major ways that you can become involved as an action researcher.

1. Use the checklist found at the end of this section to evaluate your school and teaching approaches.

2. Implement the models of excellence presented in this program. Ask yourself:
   - What outcomes do the teachers in this program accomplish that I want my students to achieve?
   - How can I find out more about their classrooms and schools?
   - Which ideas can I most easily implement in my classroom and school?
   - What will I need from my school and community?
   - How can I evaluate progress?

3. Form a team and initiate a research project. A research project can be designed to generate working solutions to a problem. The issues for your research group to address are:
   - What is the problem or question we wish to solve?
   - What will be our approach?
   - How will we assess the effectiveness of our approach?
   - What is the time frame for working on this project?
   - What resources do we have available?
   - What outcomes do we expect to achieve?

4. Investigate community needs and integrate solutions within your class activities. Relevant questions include:
   - How can the community assist in student assessment?
   - What is the community's vision of learning that defines what should be assessed?
   - How will the community benefit from improved assessment techniques?
   - What kind of relationships can my class forge with the community?

5. Establish support groups consisting of school personnel and community members. The goals of these groups are to:
   - Share teaching and learning experiences both in and out of school
   - Discuss research and theory related to learning
   - Act as mentors and coaches for one another
   - Connect goals of the community with goals of the school

Activities for Schools, Parents, and Community Members

The following are activities that groups such as your PTA, church, and local Chamber of Commerce can do together with your schools.

1. Visit your school informally for discussions using the checklist on page 23.

2. Consider ways that schools and community members can work together to provide:
   - Materials for a rich learning environment (e.g., real literature in print and audio form, computers) in which authentic assessment can occur
   - Opportunities for students and teachers to learn out of school
   - Opportunities for students to access adults as role models, tutors, aides, and experts as they engage in self assessment
   - Opportunities for students to provide community services such as surveys, newsletters, plays, and tutoring
   - Opportunities for students to participate in community affairs
   - Opportunities for administrators, teachers, or students to visit managers and company executives to learn how assessment occurs in the 'real' world

3. Promote school and community forums to debate the national education goals:
   - Invite your local television and radio stations to host school and community forums.
   - Have "revolving school/community breakfasts" (community members visit schools for breakfast once or twice a month, changing the staff and community members each time).
   - Gather information on the national education goals and their assessment.
   - Gather information on alternative models of schooling.
   - Gather information on best practices and research in the classroom.
   - Examine the vision of performance assessment set forth by CRESST.
   - Invite teachers who are involved in performance-based assessment in your area to discuss the changes being recommended by CRESST.

Some of the important questions and issues to discuss in your forums are:

- Have we reviewed the national education goals documents to arrive at a common understanding of each goal?
- What will students be like who learn in schools that achieve the goals?
- What must schools be like to achieve the goals?
- Do we agree with the goals, and how high do we rate each?
- What is the reason for national pessimism about their achievement?
- How are our schools doing now in terms of achieving each?
- Why is it important for us to achieve the goals?
- What are the consequences for our community if we don't achieve them?
- What assumptions are we making about the future in terms of Knowledge, Technology and Science, Humanities, Family, Change, Population, Minority Groups, Ecology, Jobs, Global Society, and Social Responsibility? Discuss in terms of each of the goal areas.

4. Consider ways to use this program guidebook, *Alternatives for Measuring Performance,* to promote understanding and commitment from school staff, parents, and community members.

Checklist for Excellence in Assessment

The items below are based on best practices of the teachers and researchers in Program 4. The checklist can be used to look at current practices in your school and to jointly set new goals with parents and community groups.

Vision of Learning

- ☐ Meaningful learning experiences for students and school staff
- ☐ Students encouraged to make decisions about their learning and to assess their own performance
- ☐ Restructuring to promote learning in and out of school
- ☐ High expectations for learning for all students
- ☐ A community of problem solvers in the classroom and in the school
- ☐ Teachers and administrators committed to achieving the national education goals

Curriculum and Instruction

- ☐ Identification of core concepts
- ☐ Curriculum that calls for a comprehensive repertoire of learning and assessment strategies
- ☐ Collaborative teaching and learning involving student-generated questioning and sustained dialogue among students and between students and teachers
- ☐ Teachers assessing to build new information on student strengths
- ☐ Authentic tasks in the classroom such as solving everyday problems, collecting and analyzing data, investigating patterns, and keeping journals
- ☐ Opportunities for students to engage in learning and assessment out of school with community members
- ☐ Homework that is challenging enough to be interesting but not so difficult as to cause failure
- ☐ Assessments that respect multiple cultures and perspectives
- ☐ A rich learning environment with places for children to engage in sustained problem solving and self assessment
- ☐ Instruction that enables children to develop an understanding of the purposes and methods of assessment
- ☐ Opportunities for children to decide performance criteria and method

Assessment and Grouping

- ☐ Assessment that informs and is integral to instruction
- ☐ Assessment sessions that involve the teacher, student, and parents
- ☐ Performance-based assessment such as portfolios that include drafts, journals, projects, and problem-solving logs
- ☐ Multiple opportunities to be involved in heterogeneous groupings, especially for students at risk
- ☐ Public displays of student work and rewards
- ☐ Methods of assessing group performance
- ☐ Group assessments of teacher, class, and school

Staff Development

- ☐ Opportunities for teachers to attend conferences and meetings on assessment
- ☐ Teachers as researchers, working on research projects
- ☐ Teacher or school partnerships/projects with colleges and universities
- ☐ Opportunities for teachers to observe and coach other teachers
- ☐ Opportunities for teachers to try new practices in a risk-free environment

Involvement of the Community

- ☐ Community members' and parents' participation in assessing performance as experts, aides, guides, or tutors
- ☐ Active involvement of community members on task forces for curriculum, staff development, assessment, and other areas vital to learning
- ☐ Opportunities for teachers and other school staff to visit informally with community members to discuss the life of the school, resources, and greater involvement of the community

Policies for Students at Risk

- ☐ Students at risk integrated into the social and academic life of the school
- ☐ Policies/practices to display respect for multiple cultures and role models
- ☐ A culture of fair assessment practices

SECTION 5

Biographical Information

Videoconference Participants

Eva Baker is a professor of Educational Psychology and Social Research Methods at the UCLA Graduate School of Education. She is also Director of the Center for the Study of Evaluation and Co-Director of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Dr. Baker's work in evaluation has emphasized formative evaluation for the improvement of educational programs, but more recently has focused on policy uses of assessment information. She has conducted research for or advised numerous local, state, and national agencies. Dr. Baker has been a consultant to ministries and universities in France, Japan, Italy, China, Australia, Great Britain, Qatar, and Mexico, and to organizations such as OECD and NATO. She is a congressionally appointed member of the National Council on Education Standards and Testing, and she also serves as a principal investigator on a five-year contract to assess the validity of the National Assessment of Educational Progress (NAEP).

Edmund Gordon is a member of the permanent faculty at Yale University where he holds a primary appointment as a John M. Musser Professor of Psychology. He also holds secondary appointments as Professor in the Program of African and African-American Studies; Professor, Institution of Social and Policy Studies; Professor of Child Psychology, Child Study Center; and Professor of Epidemiology and Public Health. Professor Gordon received his elementary and secondary education in segregated public schools in Goldsboro, North Carolina. He has received degrees from Howard University, Howard's Graduate School of Divinity, the American University, and Columbia University. His distinguished career includes scholarly life as a minister, clinical and counseling psychologist, research scientist, author, editor, and professor. Dr. Gordon is best known for his research on diverse human characteristics and pedagogy and the education of low status populations. He is currently working on his latest text, *Defiers of Negative Prediction: Success Against the Odds*, which is concerned with the career development of black men and women who have overcome enormous odds to become high achievers.

Charlotte Higuchi is a 3rd and 4th grade teacher at Farmdale Elementary School in Los Angeles, California. She is involved in a longitudinal study, funded by a $30,000 grant from Hewlett Packard. The study focuses on an ungraded primary school which strongly encourages parental involvement. Fifteen of Dr. Higuchi's students have been with her all four years. Her students are bilingual, ranging in ability from learning disabled to gifted. Dr. Higuchi has been awarded several teaching awards including Teacher of the Year for the California Association of Teachers of English and Southland Teacher of English. Her numerous publications include "Dialogues on the Themes of Literature: Building a Community," "Teaching Handbook" published by AFT, "Literary Resource Book for Literature," published by the California Department of Education, and an article pending with AFT entitled "Language Arts Performance-Based Assessment Grades K through 6."

Beau Fly Jones is Director of Curriculum, Instruction, and Assessment at the North Central Regional Educational Laboratory. Dr. Jones holds a Ph.D. in Educational Psychology from Northwestern University. She has expertise in the areas of cognitive learning and effective comprehension instruction. Before becoming a member of the NCREL staff, Dr. Jones served as Special Projects Coordinator for the Chicago Public Schools, and was a developer for their instructional materials in several areas, particularly social studies and reading. Since her arrival at NCREL in 1985, she has published over 25 articles and books on dimensions of thinking, cognitive instruction, and applications of technology to curriculum development. In 1990, she was Project Director for NCREL's national videoconference series, *Restructuring to Promote Learning in America's Schools.*

Judson Hixson is Director of Professional Preparation and Development at the North Central Regional Educational Laboratory and Associate Director of Technology and Information at the Midwest Regional Center for Drug-Free Schools and Communities. Mr. Hixson is a Ph.D. candidate at the University of Chicago and holds both a Bachelor and Master of Arts degree in Educational Psychology from the University of Chicago. Before joining NCREL, Mr. Hixson was Administrative Coordinator in the Chicago Public Schools. He also served as Educational Director for the Chicago Urban League, and as on-loan consultant to the National Urban League for educational policy development and congressional relations. He has teaching experience at all grade levels, as well as experience in university administration. He has served as consultant to a wide variety of organizations, including the U.S. Department of Justice, Community Relations Division.

Patrick Sweeney is a resident of the Verona Area School District and has a daughter who attends 2nd grade at Stoner Prairie Elementary School. He has served in a number of roles on school committees, including the Flexible Pacing Committee, the Authentic Assessment Action Team, and the Staff Development Retreats. Each of these groups contributed significantly toward the district's restructuring efforts. Mr. Sweeney has been involved in many aspects of education as a high school English teacher, a teaching assistant at the University of Wisconsin-Madison, and a budget and policy analyst at the Wisconsin Department of Public Instruction.

Diana Varner is the Learning Resource Coordinator at Verona Area High School in Verona, Wisconsin. Her responsibilities include assisting teachers as they make curriculum changes, tutoring students in accelerated and remedial programs, and bringing community resources into the school to provide opportunities for students to apply their knowledge in real world settings. For example, chemistry students are given the opportunity to work in the laboratory of the local hospital. Ms. Varner was a school representative for the Authentic Assessment action plan for the district.

SECTION 6

Important Assessment Resources

Video

Northwest Regional Educational Laboratory Videotapes on Assessment (ten videotapes available separately or as a set of ten). Contact IOX Assessment Associates, 5420 McConnell Ave., Los Angeles, CA 90066, (213) 822-3275.

Books

Airasian, P. W. (1991). Classroom Assessment. New York: McGraw-Hill. Easy-to-read book on basic assessment theory and practice. Includes information on performance-based assessments that will be of special interest to teachers.

Perrone, V. (Ed.). (1991). Expanding Student Assessment. Alexandria, VA: Association for Supervision and Curriculum Development. Major emphasis on better ways to assess student learning.

Wittrock, M. C., & Baker, E. L. (Eds.). (1991). Testing and Cognition. Englewood Cliffs, NJ: Prentice Hall. Series of articles from distinguished assessment researchers focusing on current assessment issues and the learning process.

Reports

The following reports on assessment are available from the National Center for Research on Evaluation, Standards, and Student Testing (CRESST). Contact Kim Hurst (310) 206-1512 to order reports. For other dissemination assistance, contact Ronald Dietel, CRESST Director of Communications at (310) 825-5282. Or write to: CRESST/UCLA Graduate School of Education, 405 Hilgard Avenue, Los Angeles, CA 90024-1522.

Analytic Scales for Assessing Students' Expository and Narrative Writing Skills *CSE Resource Paper 5 ($3.00).* These rating scales have been developed to meet the need for sound, instructionally relevant methods for assessing students' writing competence. Knowledge of students' performance utilizing these scales can provide valuable information for assessing achievement and facilitating instructional planning at the classroom, school, and district levels.

Effects of Standardized Testing on Teachers and Learning - Another Look

CSE Technical Report 334, 1991 ($5.50). This study asks "what are the effects of standardized testing on schools, and the teaching and learning processes within them? Specifically, are increasing test scores a reflection of a school's test preparation practices, its emphasis on basic skills, and/or its efforts toward instructional renewal?" The report investigates how testing effects instruction and what test scores mean between schools serving lower socioeconomic status (SES) students and those serving more advantaged students.

Complex, Performance-Based Assessments: Expectations and Validation Criteria

CSE Technical Report 331, 1991 ($3.00). This report suggests that just because performance-based measures are derived from actual student performance, it is wrong to assume that such measures are any more indicative of student achievement than are multiple-choice tests. Therefore performance-based measures need to meet validity criteria. The authors recommend eight important criteria by which to judge the validity of new students assessments, including open-ended problem-solving, portfolio assessments, and computer simulations.

Guidelines For Effective Score Reporting

CSE Technical Report 326, 1991 ($3.00). According to this report: "Despite growing national interest in testing and assessment, many states present their data in such a dry, uninteresting way that they fail to generate any attention." The authors examined current practices and recommendations of state reporting of assessment results, answering the question "how can we be sure that the vast collection of test information is properly collected, analyzed, and implemented?" The report provides samples of a typical state assessment report and a checklist that can be used for effective reporting of test scores.

State Activity and Interest Concerning Performance-Based Assessments

CSE Technical Report 322, 1990 ($2.50). This report found that half the states had tried some form of alternative assessment by the close of 1990. Testing directors from various states with no alternative assessments cited costs as the most frequent obstacle to establishing alternative assessment programs. This report suggests that getting around the financial problem may involve spreading the cost over other budgets, such as staff development and curriculum development; testing only a sample of students; sharing costs with local education agencies; or rotating the subjects tested across years.

SECTION 7

Glossary

Achievement test An examination that measures educationally relevant skills or knowledge about such subjects as reading, spelling, or mathematics.

Age norms Values representing typical or average performance of people of certain age groups.

Authentic task A task performed by students that has a high degree of similarity to tasks performed in the real world.

Average A statistic that indicates the central tendency or most typical score of a group of scores. Most often average refers to the sum of a set of scores divided by the number of scores in the set.

Battery A group of carefully selected tests that are administered to a given population, the results of which are of value individually, in combination, and totally.

Ceiling The upper limit of ability that can be measured by a particular test.

Criterion-referenced test A measurement of achievement of specific criteria or skills in terms of absolute levels of mastery. The focus is on performance of an individual as measured against a standard or criteria rather than against performance of others who take the same test, as with norm-referenced tests.

Diagnostic test An intensive, in-depth evaluation process with a relatively detailed and narrow coverage of a specific area. The purpose of this test is to determine the specific learning needs of individual students and to be able to meet those needs through regular or remedial classroom instruction.

Dimensions, traits, or subscales The sub-categories used in evaluating a performance or portfolio product (e.g., in evaluating students writing one might rate student performance on subscales such as organization, quality of content, mechanics, style).

Domain-referenced test A test in which performance is measured against a well-defined set of tasks or body of knowledge (domain). Domain-referenced tests are a specific set of criterion-referenced tests and have a similar purpose.

Grade equivalent The estimated grade level that corresponds to a given score.

Holistic scoring Scoring based upon an overall impression (as opposed to traditional test scoring which counts up specific errors and subtracts points on the basis of them). In holistic scoring the rater matches his or her overall impression to the point scale to see how the portfolio product or performance should be scored. Raters usually are directed to pay attention to particular aspects of a performance in assigning the overall score.

Informal test A non-standardized test that is designed to give an approximate index of an individual's level of ability or learning style; often teacher-constructed.

Inventory A catalog or list for assessing the absence or presence of certain attitudes, interests, behaviors, or other items regarded as relevant to a given purpose.

Item An individual question or exercise in a test or evaluative instrument.

Norm Performance standard that is established by a reference group and that describes average or typical performance. Usually norms are determined by testing a representative group and then calculating the group's test performance.

Normal curve equivalent Standard scores with a mean of 50 and a standard deviation of approximately 21.

Norm-referenced test An objective test that is standardized on a group of individuals whose performance is evaluated in relation to the performance of others; contrasted with criterion-referenced test.

Objective percent correct The percent of the items measuring a single objective that a student answers correctly.

Percentile The percent of people in the norming sample whose scores were below a given score.

Percent score The percent of items that are answered correctly.

Performance assessment An evaluation in which students are asked to engage in a complex task, often involving the creation of a product. Student performance is rated based on the process the student engages in and/or based on the product of his/her task. Many performance assessments emulate actual workplace activities or real-life skill applications that require higher order processing skills. Performance assessments can be individual or group-oriented.

Performance criteria A predetermined list of observable standards used to rate performance assessments. Effective performance criteria include considerations for validity and reliability.

Performance standards The levels of achievement pupils must reach to receive particular grades in a criterion-referenced grading system (e.g., higher than 90 receives an A, between 80 and 89 receives a B, etc.) or to be certified at particular levels of proficiency.

Portfolio A collection of representative student work over a period of time. A portfolio often documents a student's best work, and may include a variety of other kinds of process information (e.g., drafts of student work, student's self assessment of their work, parents' assessments). Portfolios may be used for evaluation of a student's abilities and improvement.

Process The intermediate steps a student takes in reaching the final performance or end-product specified by the prompt. Process includes all strategies, decisions, rough drafts, and rehearsels—whether deliberate or not—used in completing the given task.

Prompt An assignment or directions asking the student to undertake a task or series of tasks. A prompt presents the context of the situation, the problem or problems to be solved, and criteria or standards by which students will be evaluated.

Published test A test that is publicly available because it has been copyrighted and published commercially.

Rating scales A written list of performance criteria associated with a particular activity or product which an observer or rater uses to assess the pupil's performance on each criterion in terms of its quality.

Raw score The number of items that are answered correctly.

Reliability The extent to which a test is dependable, stable, and consistent when administered to the same individuals on different occasions. Technically, this is a statistical term that defines the extent to which errors of measurement are absent from a measurement instrument.

Rubric A set of guidelines for giving scores. A typical rubric states all the dimensions being assessed, contains a scale, and helps the rater place the given work properly on the scale.

Screening A fast, efficient measurement for a large population to identify individuals who may deviate in a specified area, such as the incidence of maladjustment or readiness for academic work.

Specimen set A sample set of testing materials that is available from a commercial test publisher. The sample may include a complete individual test without multiple copies or a copy of the basic test and administration procedures.

Standardized test A form of measurement that has been normed against a specific population. Standardization is obtained by administering the test to a given population and then calculating means, standard deviations, standardized scores, and percentiles. Equivalent scores are then produced for comparisons of an individual score to the norm group's performance.

Standard scores A score that is expressed as a deviation from a population mean.

Stanine One of the steps in a nine-point scale of standard scores.

Task A goal-directed assessment activity, demanding that the student use their background of knowledge and skill in a continuous way to solve a complex problem or question.

Validity The extent to which a test measures what it was intended to measure. Validity indicates the degree of accuracy of either predictions or inferences based upon a test score.

SECTION 8

References

Airaisian, P.W. (1991). Classroom assessment. New York: McGraw-Hill.

Baker, E. L. (1991, September). Alternative assessment and national education policy. Paper presented at the symposium on Limited English Proficient Students, Washington, D.C.

Charles, R., Lester, F., & O'Daffer, P. (1987). How to evaluate progress in problem solving. Reston, VA: National Council of Teachers of Mathematics.

Daves, C.W. (Ed.). (1984). The uses and misuses of tests. San Fransisco: Jossey-Bass.

Diamond, E. E., & Tuttle, C. K. (1985). Sex equity in testing. In S. S. Klein (Ed.), Handbook for achieving sex equity through education. Baltimore, MD: Johns Hopkins University Press.

Ebel, R. L., & Frisbie, D. A. (1991). Essentials of educational measurement. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Flavell, J. H. (1985). Cognitive development. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Gardner, H. (1987, May). Developing the spectrum of human intelligence. Harvard Educational Review, 76-82.

Gould, S. J. (1981). The mismeasure of man. New York: Norton and Co.

Illinois State Board of Education. (1988). An assessment handbook for Illinois schools. Springfield, IL: Author.

Lindheim, E., Morris, L.L., & Fitz-Gibbon, C.T. (1987). How to measure performance and use tests. Newbury Park, CA: Sage Publications.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. (SCE Report 331). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Nenerson, M. E., Morris, L. L., & Fitz-Gibbon, C. T. (1987). How to measure attitudes. Newbury Park, CA: Sage Publications.

Paris, S. G., Lawton, T. A., Turner, J., & Roth, J. (1991). A developmental perspective on standardized achievement testing. Educational Researcher, 20 (5).

Piaget, J. (1952). The origins of intelligence in children. New York: Norton and Co.

Popham, W.J. (1981). Modern educational measurement. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Myers, M. (1985). The teacher researcher: How to study writing in the classroom. Urbana, Illinois: National Council of Teachers of English.

Smith, M. L. (1991). Put to the test: The effects of external testing on teachers. Educational Researcher, 20 (5).

Stiggins, R. J. (1991, March). Assessment literacy. Phi Delta Kappan. 72 (7).

Tierney, R.J., Carter, M.A., & Desai, L.E. (1991). Portfolio assessment in the reading-writing classroom. Norwood, MA: Christopher-Gordon Publishers.

Wiggins, G. (May 1989). A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, 703-713.

Wise, S. L., & Plake, B. S., & Mitchell, J. V. (Eds.). (1988). Applied measurement in education. Hillsdale, NJ: Erlbaum.

Wolf, D., Bixby, J., Glenn, J. G., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (Ed.), Review of research in education (No. 17). Washington, D.C.: American Educational Research Association.

Product Order Form

Qty.	Price	Order #	Title/Author
____	$300	RES 091	**Restructuring to Promote Learning in America's Schools** (1990). A series of nine, 2-hour videotapes and guidebooks. Produced by NCREL and PBS. ■ The New Definition of Learning: The First Step for School Reform ■ The Thinking Curriculum ■ The Collaborative Classroom: Reconnecting Teachers and Learners ■ Multidimensional Assessment: Strategies for the Classroom ■ Schools as Learning Communities ■ Many Roads to Fundamental Reform: Getting Started ■ Many Roads to Fundamental Reform: Continuing the Journey ■ The Meaning of Professional Development in the 21st Century ■ Reconnecting Students at Risk to the Learning Process Individual videotape/guidebook sets are available for $35. Please specify which videoconference you want.
____	$10	HST 703	**Dimensions of Thinking: A Framework for Curriculum and Instruction** (1988). Robert J. Marzano, Ronald S. Brandt, Carolyn Sue Hughes, Beau Fly Jones, Barbara Z. Presseisen, Stuart C. Rankin, and Charles Suhor.
____	$10	HST 705	**Strategic Teaching and Learning: Cognitive Instruction in the Content Areas** (1987). Beau Fly Jones, Annemarie Sullivan Palincsar, Donna S. Ogle, and Eileen Glynn Carr.
____	$6	TC 901	**Managing Instruction for Equity and Excellence** (Facilitator's Manual) (1989). Beau Fly Jones (Ed.).
____	Free	VTC 101	**Video Teleconferencing: A Potentially Powerful Vehicle for Staff Development** (1987). Article by Beau Fly Jones.

To order these products, send a check or purchase order to:

Publications Department
North Central Regional Educational Laboratory
1900 Spring Road, Suite 300
Oak Brook, IL 60521

To receive a FREE catalog listing all NCREL products, write to the above address or call Cheryl May at 708/571-4700.